Intramolecular symmetry-adapted perturbation theory with a single-determinant wavefunction

Ewa Pastorczak,[1, a)] Antonio Prlj,[1, a)] Jérôme F. Gonthier,[2] and Clémence Corminboeuf[1, b)]

[1)] *Laboratory for Computational Molecular Design, Institut des Sciences et Ingénierie Chimiques, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland*

[2)] *Center for Computational Molecular Science and Technology, School of Chemistry and Biochemistry, School of Computational Science and Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0400, USA*

We introduce an intramolecular energy decomposition scheme for analyzing non-covalent interactions within molecules in the spirit of symmetry-adapted perturbation theory (SAPT). The proposed intra-SAPT approach is based upon the Chemical Hamiltonian of Mayer [Int. J. Quant. Chem. 23, 341 (1983)] and the recently introduced zeroth-order wavefunction [J. Chem. Phys. 140(15), 154107 (2014)]. The scheme decomposes the interaction energy between weakly-bound fragments located within the same molecule into physically meaningful components, *i.e.*, electrostatic-exchange, induction and dispersion. Here we discuss the key steps of the approach and demonstrate that a single-determinant wavefunction can already deliver a detailed and insightful description of a wide range of intramolecular non-covalent phenomena such as hydrogen bonds, dihydrogen contacts, $\pi-\pi$ stacking interactions. Intra-SAPT is also used to shed the light on competing intra- and intermolecular interactions.

[a)] E.P. and A.P contributed equally to this work

[b)] Electronic mail: clemence.corminboeuf@epfl.ch

I. INTRODUCTION

Aside from the strong covalent and ionic bonding, there exists a plethora of powerful interactions occurring between atoms and molecules. These non-covalent interactions include hydrogen[1] and halogen[2] bonds, dipole-dipole interactions, charge transfer, $\pi - \pi$ stacking, dative bonds,[3] agostic interactions,[4] as well as cation-π[5] and anion-π[6] interactions and many more.[7] Even if those interactions are more frequently associated with intermolecular complexes, their role within molecules is equally crucial, as illustrated by their impact on catalytic processes,[8] reaction barrier heights,[9,10] molecular geometries[11] or protein tertiary structures,[12] to name a few.

These interactions can be probed based on experiments[13–17] but computational techniques have played an increasingly important role over the last two decades. Those are essentially divided into two categories: the approaches that are primarily qualitative and reveal the presence of an interaction through the visualization of electron density-based functions; and methods, which provide a quantitative description of the nature of the interaction. The former category includes, for instance, the Noncovalent Interaction Index (NCI)[18] or the recent Density Regions Overlap Indicator (DORI).[19] A unique approach, combining both quantitative and qualitative features, is Bader's Quantum Theory of Atoms in Molecules (QTAIM),[20] which employs topological analysis of the electron density to reveal the existence and gain some insight into the nature of non-covalent interactions. Alternatively, various quantitative approaches have been developed to decompose the total interaction energy between molecules into physically meaningful components. Among these "energy decomposition analysis" (EDA) schemes, the most prominent are: the Kitaura-Morokuma scheme,[21] the local Møller-Plesset perturbation theory (LMP2)[22] and other linear-scaling fragment approaches.[23–25] The interaction energy terms can also be extracted through relaxing the strictly localized molecular orbitals in a field of other molecules (e.g. BLW-EDA[26], ALMO[27]). Symmetry Adapted Perturbation Theory (SAPT)[28] is a highly popular alternative, in which the interaction between monomers is introduced as a perturbation and the components of this perturbation are interpreted as electrostatics, exchange, induction and dispersion contributions. There exist different variants and implementations of SAPT including the highly accurate and computationally efficient version in PSI4,[29,30] enabling the treatment of fairly large systems, such as host-guest complexes involving DNA[31] or carbon

nanotubes.[32] A clear advantage of SAPT is the easy interpretation of the results and its firm theoretical ground.[28] Recently, Parrish and Sherrill developed a more fine-grained approach to partition the energy components into pairwise contributions from atoms or functional groups (ASAPT/FSAPT) and to visualize the results.[33,34]

Unfortunately, none of the above approaches are ideally suited for analyzing the more subtle non-covalent intramolecular interactions, although the existence of such a method would be highly valuable. The preliminary efforts to fill this gap and to expand the field of applicability of EDA schemes to a single molecule was recently accomplished by two of us with the derivation of a zeroth-order wavefunction,[35] a necessary first step towards intramolecular SAPT (as seen later, it is this wavefunction that is exploited in the present implementation). A very practical alternative, closer to standard intermolecular SAPT methods, has recently been introduced by Parrish *et al.*,[36]. This ISAPT method is built upon the functional-group SAPT[34] and adapts ideas from density matrix embedding to select the interactions and to build a zeroth-order wavefunction. The latter is then used directly in the conventional intermolecular SAPT expression. As will be seen, the method introduced here is a genuine intramolecular version of SAPT based on a novel set of expressions that makes use of the previously introduced zeroth-order expression.[35] This preliminary work has also motivated the combination of fragmentations schemes with a generalized Kohn-Sham based EDA[37] scheme that enable the analysis of intramolecular interactions, OH-π, and cation-π bonding. With the growing realization that non-covalent interactions play a significant role,[38,39] even in medium-sized molecules,[40] there is little doubt that an expanded arsenal of methods and strategies to analyze them will continue to emerge.

The paper is organized as follows: in Sec. II, we invoke the zeroth-order wavefunction of Gonthier and Corminboeuf[35] and construct a perturbation theory for intramolecular interactions for a single Slater determinant case (SD). In Sec. III, we describe the computational procedure and apply the new method to a set of illustrative molecular examples and validate its performance on intramolecular dihydrogen contacts, hydrogen bonds, $\pi - \pi$ interactions and a positively charged host-guest complex. In Sec. IV, we discuss the abilities and limitations of the proposed method as well as the perspectives for future improvement.

II. THEORY

A. The zeroth-order energy

The idea behind intramolecular SAPT (intra-SAPT) is analogous to the one of its intermolecular counterpart. First, the system is divided into fragments, the interaction between the fragments is then removed and subsequently brought back as a perturbation of the Hamiltonian. In comparison with the intermolecular scheme, the main difficulty here lies in the fact that the interacting fragments in question are not distinct monomers but selected regions within the same molecule. In quantum chemistry frameworks, all the electrons forming a molecule are described by a single wavefunction and therefore the electrons cannot be attributed to a particular atom. The electronic and nuclear partitioning is provided by Mayer's Chemical Hamiltonian approach (CHA).[41] The CHA makes use of atom-centered basis sets, and interprets the products of the interaction operators and the one- and two-electron integral kets as the physical interactions, while the bras' role is projecting those interactions onto the basis set.

The Chemical Hamiltonian approach has been originally devised to correct for the basis set superposition error (BSSE) when computing intermolecular interactions.[42–44] In Ref. 35, we showed that the same approach can be employed to probe intramolecular interactions.

As a first step toward devising a SAPT-based intramolecular energy decomposition scheme, a system (*i.e.*, a molecule or a complex) is partitioned into three fragments (note that the current implementation is limited to a three-fragment partitioning, see Fig. 1), where the interaction of interest occurs between fragments A and B, with C acting as a linker. Fragment C is generally covalently bound to both A and B but non-covalently bound fragments can be considered as well. While the nuclear partitioning associated with fragments A, B and C is straightforward, the trickier electronic partitioning is carried out through localizing subsets of electrons within each fragments using strictly localized orbitals (SLO)[45] (also known under different terminology, see Refs. 27,46–50), which by definition have non-zero coefficients only on a small number of basis functions. In practice, our implementation proceeds as follows:

1. A Hartree-Fock (HF) computation is performed on the entire system.

2. The canonical HF orbitals are projected on fragments A, B or C and then Löwdin-

orthogonalized to obtain an appropriate set of guess orbitals.

3. The guess is used to build the Fock matrix $\mathbf{F}_0$ where interactions between A and B are eliminated according to following rules:

   (a) The integrals where the product of the ket and the operator directly represents an interaction between A and B are deleted.

   (b) For integrals representing interactions within fragment A (or fragment B), the bra basis functions on fragment B (or respectively, fragment A) are projected out.

4. To ensure that orbital locality is maintained, the Fock matrix is projected on fragment X (X = A, B or C) using Stoll's algorithm[45] to get $\mathbf{F}_X^{proj}$.

5. The eigenequation $\mathbf{F}_X^{proj}\mathbf{S}\mathbf{C}_X = \epsilon_X\mathbf{S}\mathbf{C}_X$ is solved self-consistently for the orbitals $\mathbf{C}_X$ with overlap matrix $\mathbf{S}$.

The eigenequation for the orbitals $\mathbf{C}_X$ is solved under the constraint of strict orbital localization by employing Stoll's algorithm,[45] as described in more detail in Ref. 35. At convergence, one obtains occupied and virtual orbitals strictly localized on one fragment and their associated energies.

As shown in Appendix A, the zeroth-order wavefunction obtained through the above equations is in fact the right eigenvector of the zeroth-order Hamiltonian $\widehat{H}_0$:

$$\widehat{H}_0 \left|\psi_0^{(0)}\right\rangle = E_0 \left|\psi_0^{(0)}\right\rangle \tag{1}$$

$\widehat{H}_0$ is a non-Hermitian operator that can be written in closed form in second quantization (see Appendix A). As a consequence, it possesses a left eigenvector $\tilde{\psi}_0^{(0)}$ used to rewrite the above equation:

$$E_0 = \left\langle\tilde{\psi}_0^{(0)}\right| \widehat{H}_0 \left|\psi_0^{(0)}\right\rangle \tag{2}$$

The minimization of E_0 subject to the constraint of orbital localization yields the orbital optimization equations introduced above. Thus, our method is readily amenable to perturbation theory by examining the difference between the full Hamiltonian and $\widehat{H}_0$, which we do in the next section. The obtained perturbation theory formulae are expressed in terms of the occupied and virtual orbitals $\mathbf{C}_X$ and their energies.

B. Energy Decomposition

Since both the zeroth-order Hamiltonian and the perturbation are non-hermitian operators, the intramolecular perturbation theory is more easily formulated within a biorthogonal framework.[42,51]

Then, the unperturbed Hamiltonian $\widehat{H}_0$ corresponding to the 0-th order energy $E^{(0)}$ has different right and left eigenvectors. The right eigenvectors of $\widehat{H}_0$, $\left\{\left|\psi_J^{(0)}\right\rangle\right\}$, are not orthonormal to each other, but are orthogonal to the left eigenvectors, $\left\{\left\langle\tilde{\psi}_K^{(0)}\right|\right\}$ *i.e.*,

$$\left\langle\tilde{\psi}_K^{(0)}|\psi_J^{(0)}\right\rangle = \delta_{KJ} \quad . \tag{3}$$

The perturbation in the biorthogonal formulation (and second quantization notation) is expressed as a sum

$$\widehat{H} - \widehat{H}^{(0)} = \widehat{W}_{AB} + \widehat{H}_{BE} \quad , \tag{4}$$

where the first component on the right takes the following form

$$\widehat{W}_{AB} = \sum_{i\in A}\sum_{k}\left\langle\tilde{k}\left|\hat{V}_B\right|i\right\rangle k^+\tilde{i}^- + \sum_{i\in B}\sum_{k}\left\langle\tilde{k}\left|\hat{V}_A\right|i\right\rangle k^+\tilde{i}^- + \sum_{k\in A}\sum_{l\in B}\sum_{ij}\left\langle\tilde{i}\tilde{j}||kl\right\rangle i^+j^+\tilde{l}^-\tilde{k}^- \tag{5}$$

where $\hat{V}_A$, $\hat{V}_B$ are the electrostatic potentials of the nuclei of fragments A and B, respectively, and the biorthogonal spinorbitals $\left\langle\tilde{i}\right|$ are defined by the relation

$$\left\langle\tilde{i}|j\right\rangle = \delta_{ij} \quad , \tag{6}$$

while k^+, $\tilde{i}^-$ are, respectively, the non-hermitian covariant creation and contravariant annihilation operators.[52] Please note that in the above and the following equations where no other indication is given, the indices run over the entire set of orbitals.

The remaining component of the perturbation, $\widehat{H}_{BE}$, is associated with basis set effects and does not contribute to the physical part of the AB interaction (see Appendix B). It can thus be neglected.

The first-order energy correction (including also the classical internuclear repulsion), $E^{(1)} = \left\langle\tilde{\Psi}^{(0)}\left|\widehat{W}_{AB}\right|\Psi^{(0)}\right\rangle + \frac{1}{2}\sum_{I\in A,J\in B}\frac{Z_IZ_J}{R_{IJ}}$ then takes the form

$$E^{(1)} = \sum_{i\in A}^{occ}\left\langle\tilde{i}\left|\hat{V}_B\right|i\right\rangle + \sum_{i\in B}^{occ}\left\langle\tilde{i}\left|\hat{V}_A\right|i\right\rangle + \sum_{k\in A}^{occ}\sum_{l\in B}^{occ}\left\langle\tilde{k}\tilde{l}||kl\right\rangle + \sum_{I\in A,J\in B}\frac{Z_IZ_J}{R_{IJ}} \quad . \tag{7}$$

The first order correction (7) corresponds to the sum of the Coulomb and exchange interactions between the fragments, although the exchange component cannot be isolated in the biorthogonal formulation.

The second-order correction takes the form

$$E^{(2)} = -\sum_{exc} \frac{\left\langle \tilde{\Psi}_{exc} \middle| \widehat{W}_{AB} \middle| \Psi^{(0)} \right\rangle \left\langle \tilde{\Psi}^{(0)} \middle| \widehat{W}_{AB} \middle| \Psi_{exc} \right\rangle}{E^{(0)}_{exc} - E^{(0)}_{0}} \tag{8}$$

where $\sum_{exc}$ denotes a summation over all the excited determinants, and $E^{(0)}_{exc}$ are the zeroth-order energies of the excited states. It can be expressed as a sum of three components (for an extended description see Appendix C).

$$E^{(2)} = E_{pol} + E_{deloc} + E_{disp} \quad , \tag{9}$$

where

$$\begin{aligned} E_{pol} = &-\sum_{a\in A}^{occ}\sum_{b\in A}^{virt} \frac{\left(\left\langle \tilde{a} \middle| \hat{V}_B \middle| b \right\rangle + \sum_{l\in B}^{occ} \left\langle \tilde{a}\tilde{l}||bl \right\rangle\right)\left(\left\langle \tilde{b} \middle| \hat{V}_B \middle| a \right\rangle + \sum_{l\in B}^{occ} \left\langle \tilde{b}\tilde{l}||al \right\rangle\right)}{\varepsilon_b - \varepsilon_a} \\ &-\sum_{a\in B}^{occ}\sum_{b\in B}^{virt} \frac{\left(\left\langle \tilde{a} \middle| \hat{V}_A \middle| b \right\rangle + \sum_{l\in A}^{occ} \left\langle \tilde{a}\tilde{l}||bl \right\rangle\right)\left(\left\langle \tilde{b} \middle| \hat{V}_A \middle| a \right\rangle + \sum_{l\in A}^{occ} \left\langle \tilde{b}\tilde{l}||al \right\rangle\right)}{\varepsilon_b - \varepsilon_a} \end{aligned} \tag{10}$$

corresponds to the polarization energy,

$$\begin{aligned} E_{deloc} = &-\sum_{a\in A}^{occ}\sum_{b\in B}^{virt} \frac{\left(\left\langle \tilde{a} \middle| \hat{V}_A \middle| b \right\rangle + \sum_{l\in A}^{occ} \left\langle \tilde{a}\tilde{l}||bl \right\rangle\right)\left(\left\langle \tilde{b} \middle| \hat{V}_B \middle| a \right\rangle + \sum_{l\in B}^{occ} \left\langle \tilde{b}\tilde{l}||al \right\rangle\right)}{\varepsilon_b - \varepsilon_a} \\ &-\sum_{a\in B}^{occ}\sum_{b\in A}^{virt} \frac{\left(\left\langle \tilde{a} \middle| \hat{V}_B \middle| b \right\rangle + \sum_{l\in B}^{occ} \left\langle \tilde{a}\tilde{l}||bl \right\rangle\right)\left(\left\langle \tilde{b} \middle| \hat{V}_A \middle| a \right\rangle + \sum_{l\in A}^{occ} \left\langle \tilde{b}\tilde{l}||al \right\rangle\right)}{\varepsilon_b - \varepsilon_a} \end{aligned} \tag{11}$$

to the delocalization or charge-transfer energy and

$$E_{disp} = -\sum_{a\in A}^{occ}\sum_{c\in B}^{occ}\sum_{b\in A}^{virt}\sum_{d\in B}^{virt} \frac{\langle \tilde{a}\tilde{c}||bd\rangle \left\langle \tilde{b}\tilde{d}||ac \right\rangle}{\varepsilon_d - \varepsilon_c + \varepsilon_b - \varepsilon_a} - \sum_{a\in A}^{occ}\sum_{c\in B}^{occ}\sum_{b\in B}^{virt}\sum_{d\in A}^{virt} \frac{\langle \tilde{a}\tilde{c}||bd\rangle \left\langle \tilde{b}\tilde{d}||ac \right\rangle}{\varepsilon_d - \varepsilon_c + \varepsilon_b - \varepsilon_a} \tag{12}$$

to the London dispersion term. In Eqs. (10) - (12) $\{\varepsilon_x\}_{x=a,b,c,d}$ denote orbital energies associated with the zeroth-order wavefunction.

The expressions for the intramolecular interaction components (7) and (10) - (12) resemble the ones obtained by Surjan *et al.*[42] for intermolecular interactions (in fact they are

identical in cases for which the linker is absent), but in the intra-SAPT formulas the presence of the middle fragment manifests itself through the orbitals and their energies. In the single-determinant approximation, the three- and higher-body terms are equal to zero in the first and second-order corrections. Note also that, in principle, it is possible to introduce higher-order correction (although it is very cumbersome, see Ref. 53). Without these, the convergence of the perturbation series is difficult to assess. However, we expect a similar rate of convergence seen in other unrestricted perturbation theory variants, e.g., UMP2. The convergence of unrestricted methods is generally poorer than in their restricted counterparts, a fact that is generally attributed to spin contamination of the wave function.[54,55] Regardless, the energy components should be less sensitive to these features (see e.g. Ref. 56) than binding energies and reaction barriers, properties on which the convergence rate is usually probed.

III. ILLUSTRATIVE EXAMPLES

The zeroth-order wavefunction and perturbative expressions introduced above leads to a unique perturbation theory-based scheme specifically tailored for decomposing non-covalent interactions within molecules.

Of course, the simplicity associated with approximating the wavefunction as a single Slater determinant is appealing but it imposes certain limitations when using intra-SAPT. A single SD wavefunction is, for instance, not suitable for systems with significant multi-configuration character. Additionally, within the SD approximation, the partitioning of the system into fragments induces a spin contamination of the zeroth-order wavefunction. The contamination will be most problematic in situations where the linker is very small, leading to fictitious interactions near the border of the fragment partitioning, such as a too attractive first-order term or even a slightly positive induction contribution.

Another limitation inherent to any PT scheme is that the perturbation representing the interaction should be small. In line with the issue associated with the spin-contamination, this condition imposes that the interacting fragments should lie fairly far from each other *i.e.*, the covalently-bound linker fragment should correspond, to more than one heavy atom.

Finally, the choice of the basis set can be rather sensitive. On one hand, it is known[57] that the SLO-based approaches lack a complete basis set limit (CBS) and that only moderate

size basis sets should be used. On the other hand, the proper description of dispersion interaction depends on the presence of sufficient number of virtual orbitals. Akin to other EDA schemes exploiting atom-centered basis sets,[58,59] double zeta (polarized) basis sets with no diffuse functions are generally recommended.

In the following sections, we provide applications of intra-SAPT for a few illustrative systems in which the previously mentioned limitations are minimal.

A. Computational details

In each of the proposed examples, the zeroth-order wavefunction was obtained as described in Ref. 35, based on an unrestricted spin formalism. The formulas were implemented in a developer version of the Molpro software package.[60]

First, the method is validated on prototypical rare gas dimer systems (see Supporting Information[61]), showing that when no linker is present, the method reduces to an intermolecular method similar to that of Surjan *et al.*[42]

Unless otherwise stated, all the computations were done with the 6-31G[62] basis set on a MP2/6-31G* geometry. The limitation of the current, developer version of the code prevented the use of larger basis sets for all of the investigated systems, but results for larger basis sets are provided whenever it was possible (see Table I in Sec. III B as well as Tables S3, S4 and Fig. S1 and S2 in Supplementary Material[61]). For the sake of clarity, the terms E_{deloc} and E_{pol} were summed and labeled as an "induction" term, E_{ind}, in most plots and tables.

In all the investigated systems, the linker (see fragment C from Fig. 1) is bonded either non-covalently or through single covalent bonds to the interacting fragments A and B. To minimize spurious ionic interactions, the occupied spinorbitals were distributed in such a way that each of the interacting fragments corresponds to an open-shell system (as opposed to closed-shell ionic fragments).

B. Hairpin alkanes - stabilizing effect of dihydrogen contacts

Structurally simple and ubiquitous in nature[63,64] unbranched alkanes are a perfect illustration of the importance of London dispersion interactions within molecules. In particular,

the question regarding at which carbon chain length the folded, "hairpin"-like, conformation is favored over its linear counterpart, has been recently addressed by both experimentalists[65] and theoreticians.[66,67] According to the current consensus, the last globally stable extended alkane is either $C_{17}H_{36}$ or $C_{18}H_{38}$. However, the reason for which the alkane molecules do fold, has not yet received a direct answer. In analogy to dimers of n-alkanes, polyhedranes[68] and spatially aligned [n]ladderanes,[69] it is assumed that the cumulated dispersion interactions arising from the parallel fragments of carbon chains are at the origin of this conformational isomeric process. In this present context, intra-SAPT can directly reveal the nature of the interactions between the carbon chains in the folded forms.

Here, we investigate carbon chains of lengths from $N = 8$ to $N = 19$. The geometries taken from Ref. 67 show three well-defined fragments (see Fig. 1) with a middle C_4H_8 fragment in each alkane, which naturally serves as the linker between the two interacting hydrocarbon chains of either equal length (when N is an even number) or differing by one CH_2 carbon atom (when N is odd).

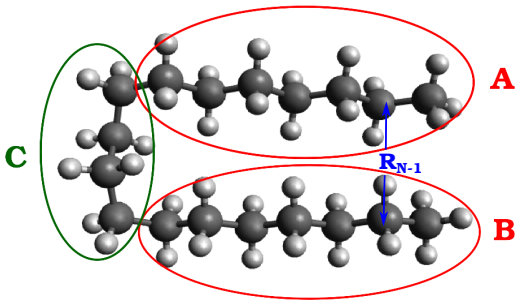

Figure 1. Backbone structure of $C_{18}H_{38}$ partitioned into three fragments. Geometries are taken from Ref. 67.

We must distinguish (see Fig. 2) between the odd- and even- alkane chains, differing by their number of carbon atoms and of hydrogen atom contacts, which potentially dictate the nature of the interactions. Actually, the overall trends for the energy terms with respect to the number of carbon atoms remain similar for both the odd and even cases. The induction term is obviously very small as both fragments are neutral, symmetric or nearly symmetric and spatially distinct. Fig.2 shows that up to $N = 12$ the fragments are short and far from each other which results in a near-zero dispersion term. At medium chain lengths, up to

$N = 16$ the increase in attractive dispersion contribution is compensated by the growth of the repulsive first-order term. This result is essentially in line with the latest theoretical assessments[67] that place $C_{16}H_{34}$ as the largest globally stable unfolded alkane.

As the energetic penalty associated with further distortion of the interfragment bond angles decreases with increasing the chain length, the fragments get closer. Dispersion interactions start to dominate from $C_{16}H_{34}$, resulting in a slightly attractive total interaction. The increase of the dispersion energy originating from the elongation of the carbon chains (*i.e.,* the increased number of interacting electron pairs) causes the side chains to approach one another even more such as to maximize the non-covalent interactions.

Considering that all the investigated geometries (see Ref. 67) were optimized with dispersion-corrected density functionals, it is clear that while for short chains the covalent interfragment bonds play a decisive role in shaping the geometry, the non-covalent interaction (originating from the interplay of Pauli repulsion, electrostatic interaction and London dispersion) becomes increasingly determinant for longer and more flexible chains. This realization is likely to be relevant for other hairpin-like structures, e.g. phospholipids and hairpin peptides.

It is worthwhile noting that the total interactions arise from a balance of relatively small energy contributions of opposite sign and that it is certainly more reasonable to analyze the trends than an energy value for a specific alkane.

The dependence on the basis set is tested through computing the interaction energies of the small hairpin alkanes ($N = 8, ..., 14$) with the 6-31G*[70] basis set (see Fig. S1 in SI[61]). The first order term is less repulsive with 6-31G*, but the trends are identical. The interaction energy components obtained for C_8H_{18}, with a wider selection of basis sets, are presented in Table I. The tested basis sets of comparable size produce similar outcome. To examine the sensitivity of intra-SAPT method to small geometry changes, supplementary computations using the geometries presented in Ref. 66 (optimized at MP2/cc-pVTZ level) were performed. The results turned out to be very similar to the ones presented in Fig.2 (less than 0.4 kcal/mol of difference in a single component). Those additional tests demonstrate the robustness of the qualitative intra-SAPT trends applied to the hairpin alkanes.

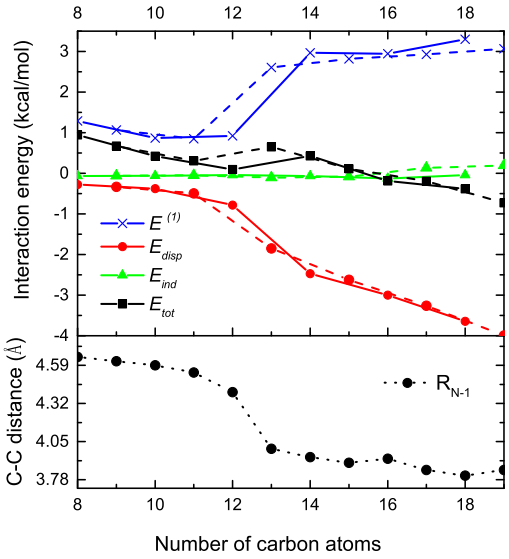

Figure 2. Interaction energy components in hairpin alkanes between fragments A and B (see Fig. 1), 6-31G basis set (upper plot) and the distances of the closest C-C pair in the opposite chains (lower plot, see Fig. 1) plotted against the number of carbons. The dashed lines are used for the odd-numbered alkanes, the continuous ones for the even-numbered alkanes. Geometries are taken from Ref. 67.

Basis set	$E^{(1)}$	E_{ind}	E_{disp}	E_{tot}
6-31G	1.29	−0.07	−0.28	0.94
6-31G*	0.98	−0.06	−0.30	0.62
6-311G	2.49	0.02	−0.33	2.18
def2-SVP[71]	2.06	0.03	−0.49	1.73
cc-pVDZ[72]	1.34	0.07	−0.32	1.09

Table I. Components of the intramolecular interaction between C_2H_5 groups in C_8H_{18} in kcal/mol.

C. $\pi - \pi$ stacking interactions

π-stacked aromatic chromophores is another appealing class of geometrical patterns that leads to unique properties.[73] Their interaction is often analyzed using simpler model systems such as benzene[74] and substituted benzene dimers.[75,76] With intra-SAPT, however, it is possible to access information regarding the same interaction occurring within a molecule. This

intramolecular framework opens the possibilities to study closer interchromophore distances that go below van der Waals radii and, which would not be possible with distinct molecules.

Here, we present two examples featuring intramolecular $\pi - \pi$ stacking: 3-phenyl-2(2-phenylacetylamino)propionic acid (Phe-L-PHA) (see Fig. 3, left) and an analogous molecule with the lower phenyl ring replaced by a perfluorophenyl ring (PFB-L-PHA, see Fig. 3, right). The latter system is employed as a typical building block in supramolecular hydrogelators,[77] triggered by the intramolecular interaction between the phenyl and the perfluorophenyl ring. Both structures were optimized at the MP2/6-31G* level. The interacting fragments, A and B in Phe-L-PHA and PFB-L-PHA are the two aromatic rings.

As illustrated in Table II, the interaction between the two phenyl rings in Phe-L-PHA is slightly repulsive, with an attractive contribution arising from dispersion and, to a lesser extent, from induction that is compensated by the repulsive first-order term. Note that the phenyl rings in this optimized geometry are rather close - the distance between the ring centers is $R = 4.201\,\text{Å}$ (see Fig. 3), but the upper ring is tilted, which results in one of the H-C distances being as close as $R_{HC} = 2.932\,\text{Å}$. This geometry is quite far from the typical $\pi - \pi$ stacking "sandwich" conformation (with the rings about $3.8\,\text{Å}$ apart) and does not correspond to any minimum or saddle point of a benzene dimer.[74] This situation is reminiscent from that of the medium-size hairpin alkanes and suggest that most medium-size apolar molecules use the attractive dispersion forces to fight against the repulsive wall and form more compact geometries in which the total attraction between fragments is fairly small. The intra-SAPT trend is fully consistent with a SAPT(HF) computation performed on a benzene dimer constrained in the geometry of the Phe-L-PHA phenyl rings using the same basis set (6-31G). SAPT(HF) also reveals a slight repulsion ($0.97\,\text{kcal/mol}$) with the first-order term of $4.33\,\text{kcal/mol}$ and the dispersion contribution of $-3.36\,\text{kcal/mol}$.

The PFB-L-PHA conformation is closer to a parallel ring arrangement, which is characterized by a larger dispersion energy term. In line with the benzene-hexafluorobenzene complex, the first-order term is less repulsive than for Phe-L-PHA. This difference has often been attributed to the opposite sign of the quadrupole moments of the phenyl ring and perfluorophenyl ring,[78] respectively, but has more recently been explained in terms of local dipole-dipole interactions between the substituents and the phenyl ring.[79] The two aforementioned effects lead to a negligible total interaction between PFB and Phe rings, which is again consistent with the SAPT(HF)/6-31G result for a phenyl-pentafluorophenyl

complex in the same configuration (the first order energy component 4.88 kcal/mol is compensated by the dispersion contribution −5.44 kcal/mol summing up to a total interaction of −0.56 kcal/mol). The small induction and induction-exchange contributions lower the interaction energy further to −0.49 kcal/mol.

The remarkable agreement of intra-SAPT and SAPT(HF) results in this case does not only validate the intra-SAPT results, but also indicates that, in this case, the linker does not significantly influence the $\pi - \pi$ interactions. Nevertheless, the linker plays a decisive role in placing the aromatic rings in an orientation that is dictated by the bond and angle strain and not by the maximization of non-covalent interactions. In this respect, the two systems presented here are similar to short hairpin alkanes, where the interaction between the side chains is slightly repulsive.

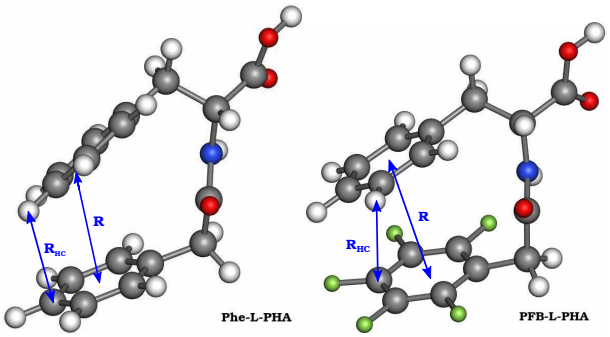

Figure 3. Backbone structures of Phe-L-PHA (left) and PFB-L-PHA (right). Color code: O - red, N - blue, F - green, C - grey, H - white. Geometries optimized at the MP2/6-31G* level. Distances between ring centers: Phe-L-PHA - $R = 4.201$ Å, PFB-L-PHA - $R = 3.268$ Å, distances between the closest C-H pair: Phe-L-PHA - $R_{HC} = 2.932$ Å, PFB-L-PHA - $R_{HC} = 3.268$ Å.

	$E^{(1)}$	E_{ind}	E_{disp}	E_{tot}
Phe-L-PHA	5.04	−0.44	−2.07	1.97
PFB-L-PHA	3.88	−0.60	−3.10	0.18

Table II. Interaction energy components in kcal/mol between the phenyl rings (Phe-L-PHA) and between the phenyl and pentafluorobenzyl ring (PFB-L-PHA). Computations at the 6-31G level.

D. Intramolecular hydrogen bonds

In its simplest picture, the physical nature of intermolecular hydrogen bonds is essentially discussed in terms of electrostatic interactions, which differs significantly from the previous examples. While the consensus is that the electrostatic contribution accounts for most of the interaction energy, EDA analysis have also highlighted the importance of contributions such as charge transfer, exchange and dispersion.[80–82] The intramolecular case is even more controversial as there is no straightforward way to establish its attractive character within a molecule (see e.g. Refs 83,84). The aminoalcohol series (*i.e.*, 2-aminoethanol, 3-aminopropanol, 4-aminobutanol and 5-aminopentanol) is a good test case giving access to different orientations and hydrogen bond distances between the hydroxyl and the amine group. In fact, the H-bond within small aminoalcohols is believed to be one of the strongest.

The geometries were optimized at the MP2/6-31G* level. The studied interactions were between the hydroxyl group, fragment A, and the NH_2 group, fragment B.

As illustrated in Fig. 4, the hydrogen bond is strongly attractive due to the electrostatic interaction, with almost no contribution from the second-order terms. The total interaction in the smallest 2-aminoethanol ($N = 4$), $E_{tot} = -6.51$ kcal/mol, is of similar strength as the one in the ammonia-water complex (-6.36 kcal/mol).[85]. The hydrogen bond distance shortens from 2.165 Å in 2-aminoethanol down to 1.851 Å in 4-aminobutanol. The bond angle is also dramatically affected going from the pseudo 4- to 6-membered ring as illustrated by the collinearization the OHN angle (from 118° to 158°). However, the position and orientation of the hydroxyl group with respect to NH_2 change very slightly between 4-aminobutanol and 5-aminopentanol ($N = 7$) as evidenced by their similar interactions. Akin to the former examples (Sec. III B) the maximization of the non-covalent interaction only occurs once sufficient flexibility is achieved within a molecule. Still, the aminoalcohol series shows that those interactions can even have a strong impact on the geometry of small systems. This is a distinctive feature from the hairpin alkane case, which originates from the different nature of the dominant non-covalent interactions. Aminoalcohols are essentially held together by stronger, electrostatic contributions that do not depend on the number of interacting electron pairs and are already efficient in small systems. In contrast, alkane chains must reach a critical length to benefit from the stabilization arising from dispersion.

The more pronounced hydrogen bond interaction in 3-aminopropanol as compared to the

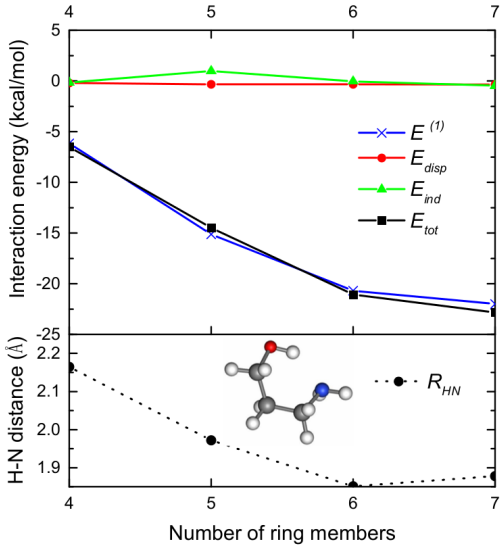

Figure 4. Interaction contributions between the hydroxyl and the amine groups in aminoalcohols (upper plot) and the H-N distances (lower plot) in the hydrogen bond, plotted against number of the ring members, 6-31G basis set. Geometries optimized at the MP2/6-31G* level.

smaller aminoalcohol is consistent with the red shift frequency measured in the vibrational gas phase spectra of the former compound.[86,87] The magnitude of the total interaction energy for the last two compounds ($E_{tot} = -21.06\,\text{kcal/mol}$ and $E_{tot} = -22.84\,\text{kcal/mol}$ for 4-aminobutanol and 5-aminopentanol, respectively) is in the range of the strongest hydrogen bonds[88] but one cannot exclude an underestimation of the Pauli repulsion owing to the fragment partitioning. The description of the interfragment bond could however be improved by introducing a spin-coupling scheme,[89] which is planned.

E. Host-guest complexes with a cationic guest

While intra-SAPT is essentially designed to decompose intramolecular interactions, it is also highly valuable for probing the competing non-covalent interaction between two functional groups belonging to the same molecule and a particular fragment such as another molecule, ion, *etc.*

Two examples of such systems are provided in Fig. 5, where a cationic lithium atom intercalated between two neutral functional groups generates complexes **1** and **2**.

For each complex two computations are performed: one to probe the interaction between the cationic lithium and the sulfur-containing fragment (e.g., thioester or thionoester, see Table III and Fig. 5) and another with Li^+ interacting with the ester functional group. In both cases the the middle CH_2 fragment was taken as the linker.

It is apparent from the data in Table III that the competition between the guest and the functional groups is larger in complex **1** than in **2**. In **1**, both the ester and thioester fragments bound strongly to the lithium with a marginal advantage to the electron-richer thioester. The decomposition via intra-SAPT indicates a large electronic redistribution within the ester-type groups engendered by the presence of the cationic lithium causing both their polarization (large E_{pol} contributions) and a large delocalization (i.e., charge transfer). The delocalization is significantly larger with the case of the richer thioester group, which is more prone to donate. The electrostatic terms are in the same range as the polarization contributions, whereas the binding contribution from dispersion is negligible in both interactions.

Intra-SAPT also indicates that **2** featuring the thionoester group is somewhat trickier and more frustrated chemically.[90] The guest can interact with both an oxygen or sulfur atom but the interaction with the ester group is clearly stronger and very similar to the interactions observed in complex **1**. To interpret the much lower binding affinity with the thionoester function and the behavior of the individual energy component, one first needs to analyze what happens in the zeroth-order energy computations. When the guest and the sulfur-containing functional group are taken as the interacting fragments A and B, respectively, the relatively strong interaction between them is removed in the zeroth-order computation. In the absence of interacting cationic lithium, the electrons within fragment B are pulled away from sulfur to the more electronegative oxygen atom, creating a dipole with the sulfur atom as the positive pole. Once the interaction is brought back by the perturbation, this polarization engenders a repulsive first-order energy component that is largely compensated by the delocalization term, associated with the ion pulling the electrons towards itself. The polarization term also brings significant attractive contribution, like in all the interactions for both complexes, but is slightly smaller.

The difference between these two complexes agrees with the distinct chemical properties of the two sulfur functional groups.[90] The presence of the charged lithium atom induces a larger perturbation (and frustration) in the thionoester case, in which the sulfur atom is

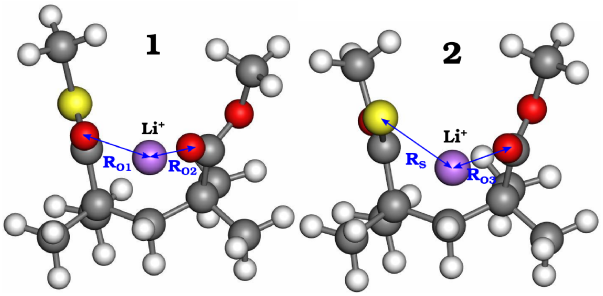

Figure 5. Backbone structures of the host-guest complexes. Color code: O - red, S - yellow, Li - purple, C - grey, H - white. Geometries optimized at the MP2/6-31G* level. Distances $R_{O1} = 1.849$ Å, $R_{O2} = 1.845$ Å, $R_{O3} = 1.835$ Å, $R_S = 2.388$ Å.

System	Fragment A	Fragment B	$E^{(1)}$	E_{deloc}	E_{pol}	E_{disp}	E_{tot}
1	Li^{+a}	O=C(R)−S·Me	−8.00	−17.11	−10.44	−0.03	−35.58
		O=C(R)−O·Me	−11.80	−10.18	−12.90	−0.04	−34.92
2		S=C(R)−O·Me	54.50	−47.28	−7.98	−0.11	−0.88
		O=C(R)−O·Me	−8.39	−18.61	−10.99	−0.03	−38.03

Table III. Interaction energy components between Li^+ and fragments of molecules A and B in kcal/mol. R=C $(Me)_2$. Computations with the 6-31G basis set. [a]Within our scheme it is the whole system positively charged rather than the single lithium atom.

positioned between the more electron-demanding oxygen atom and the cationic lithium. In contrast, the thioester oxygen atom benefits from the close proximity and electron sharing of the richer sulfur atom without compromise. In a sense, complex 2 may illustrate the limit of the applicability of a perturbation-based method: the electrostatic interaction between the the cationic lithium guest and the fragments is so large that it does affect the electron density distribution. This results in an overestimation of the electrostatic repulsion, which is then compensated by the delocalization term. While it is reasonable to interpret the total interaction energies between the cationic lithium and the thienoester group as much weaker than the other interactions, the interpretation of the individual energy terms is delicate in this limiting case.

IV. CONCLUSIONS

We introduced a unique intramolecular variant of SAPT capable of describing the nature of non-covalent intramolecular interactions. The method is clearly different from other existing methods and relies upon a new wavefunction and set of expressions specifically developed for the decomposition of non-covalent intramolecular interactions. This decomposition scheme complements the recently introduced ISAPT method of Parrish *et al.*[36] which makes use of the the standard two-body SAPT methodology via Hartree-Fock embedding.

The approach was used to decompose the interaction energy of hydrogen bonds, $\pi - \pi$ stacked rings, alkane chains within molecules. These illustrative examples along with others involving the competing interaction between a cation and different functional groups belonging to the same molecules demonstrate that intra-SAPT is able to treat both inter- and intramolecular phenomena on an equal footing.

While the present implementation suffers from certain limitations associated with the use of a wavefunction based on a single Slater determinant and the imperfect description of the interfragment bonds, the issue could be overcome through the introduction of spin-coupling scheme.[89] Akin to other EDA schemes that make use of strictly localized orbitals,[57,91] intra-SAPT remains ill-defined in the CBS limit but this should not prevent its applications and ability to uncover insightful information previously inaccessible.

In fact, the current implementation has already been able to identify how the role of dispersion evolves with elongating the carbon chain in hairpin alkanes: from being negligible in the short chains up to being the driving force for the formation of the hairpin pattern in the long chains. Similar maximization of the non-covalent interactions was observed in the aminoalcohol series.

Overall the afore-mentioned trends indicate that once a critical size is reached, an optimal molecular conformation should coincide with a local (broadly understood) van der Waals energy minimum. This phenomenon would certainly be worth further examination.

There is no doubt that both ISAPT and intra-SAPT have opened the door to a broad range of new exciting applications. The forthcoming comparisons of these methods not only will illustrate their usefulness and complementarity, but also will assist in the development of future improvements.

ACKNOWLEDGMENTS

Funding from EPFL and from Swiss NSF Grants no. "200021_137529" and "200021_156001" is gratefully acknowledged. We thank Dr. Ganna Gryn'ova for discussions on the illustrative examples.

Appendix A: Equations for orbital optimization

Here we demonstrate the connection between $\widehat{H}_0$ and the modified Fock matrix introduced in Sec. II A. We start by writing the zeroth-order Hamiltonian $\widehat{H}_0$ explicitly:

$$\begin{aligned}\widehat{H}_0 = &\sum_{i\in A}\sum_{k}\left(\hat{h}_{AC}\right)_{\tilde{k}i} k^+\tilde{i}^- + \sum_{i\in B}\sum_{k}\left(\hat{h}_{BC}\right)_{\tilde{k}i} k^+\tilde{i}^- + \sum_{i\in C}\sum_{k}\hat{h}_{\tilde{k}i}k^+\tilde{i}^- \\ &+\frac{1}{2}\sum_{k,l\in A}\sum_{i,j}(\hat{v}_{AC})^{kl}_{\tilde{i}\tilde{j}}\, i^+j^+\tilde{l}^-\tilde{k}^- + \frac{1}{2}\sum_{k,l\in B}\sum_{i,j}(\hat{v}_{BC})^{kl}_{\tilde{i}\tilde{j}}\, i^+j^+\tilde{l}^-\tilde{k}^- \\ &+\frac{1}{2}\sum_{k,l\notin\{AB\}}\sum_{i,j} v^{kl}_{\tilde{i}\tilde{j}} i^+j^+\tilde{l}^-\tilde{k}^- + V^A_{nuc} + V^B_{nuc} + \frac{1}{2}\sum_{I,J\notin AB}\frac{Z_IZ_J}{R_{IJ}} \qquad \text{(A1)}\end{aligned}$$

$$\qquad \text{(A2)}$$

where $k,l \notin \{AB\}$ means that the pair kl cannot be in AA, BB, AB or BA and $V^X_{nuc} = \frac{1}{2}\sum_{i,j\in X}\frac{Z_iZ_j}{\mathbf{R}_{ij}}$. In addition, we have introduced the following integral definitions:

$$\left(\hat{h}_{AC}\right)_{\tilde{k}i} = \left\langle \tilde{k}\left|\left(\hat{P}_{AC}\right)\left(\hat{V}_A+\hat{V}_C+\hat{T}\right)\right|i\right\rangle \qquad \text{(A3)}$$

$$\left(\hat{h}_{BC}\right)_{\tilde{k}i} = \left\langle \tilde{k}\left|\left(\hat{P}_{BC}\right)\left(\hat{V}_B+\hat{V}_C+\hat{T}\right)\right|i\right\rangle \qquad \text{(A4)}$$

$$(\hat{v}_{AC})^{kl}_{\tilde{i}\tilde{j}} = \left\langle \tilde{i}\tilde{j}\left|\left(\hat{P}_{AC}(1)\hat{P}_{AC}(2)\right)\frac{1}{r_{12}}\right|kl\right\rangle \qquad \text{(A5)}$$

$$(\hat{v}_{BC})^{kl}_{\tilde{i}\tilde{j}} = \left\langle \tilde{i}\tilde{j}\left|\left(\hat{P}_{BC}(1)\hat{P}_{BC}(2)\right)\frac{1}{r_{12}}\right|kl\right\rangle \qquad \text{(A6)}$$

where $\hat{T}$ is the kinetic energy operator, $\hat{V}_A$, $\hat{V}_B$, $\hat{V}_C$ are the electrostatic potentials of the nuclei of fragments A, B and C and $\hat{P}_{AC}$ and $\hat{P}_{BC}$ are projection operators projecting on pairs of fragments AC and BC, respectively:

$$\hat{P}_{AC} = \sum_{\mu,\nu\in AC}|\mu\rangle\left(\mathbf{S}^{(AC)}\right)^{-1}_{\mu\nu}\langle\nu| \qquad \text{(A7)}$$

and $\mathbf{S}^{(AC)}$ is the overlap matrix of the basis functions on A and C. We also have the usual one-electron potential integral $\hat{h}_{\tilde{k}i} = \left\langle \tilde{k} \left| \hat{V}_A + \hat{V}_B + \hat{V}_C \right| i \right\rangle$ and two-electron repulsion integral $v_{\tilde{i}\tilde{j}}^{kl} = \left\langle \tilde{i}\tilde{j} \left| \frac{1}{r_{12}} \right| kl \right\rangle$.

The zeroth-order energy is written in terms of $\psi_0^{(0)}$ and easily evaluated since our creation k^+ and annihilation $\tilde{i}^-$ operators obey the usual commutation and anticommutation relationships.

$$\begin{aligned} E_0 &= \left\langle \tilde{\psi}_0^{(0)} \left| \widehat{H}_0 \right| \psi_0^{(0)} \right\rangle \\ &= \sum_{i\in A,occ} \left(\hat{h}_{AC}\right)_{\tilde{i}i} + \sum_{i\in B,occ} \left(\hat{h}_{BC}\right)_{\tilde{i}i} + \sum_{i\in C,occ} \hat{h}_{\tilde{i}i} + \frac{1}{2} \sum_{i,j\in A,occ} \left((\hat{v}_{AC})_{\tilde{i}\tilde{j}}^{ij} - (\hat{v}_{AC})_{\tilde{i}\tilde{j}}^{ji} \right) \\ &\quad + \frac{1}{2} \sum_{i,j\in B,occ} \left((\hat{v}_{BC})_{\tilde{i}\tilde{j}}^{ij} - (\hat{v}_{BC})_{\tilde{i}\tilde{j}}^{ji} \right) + \frac{1}{2} \sum_{k,l\notin\{AB\}} \left(v_{\tilde{i}\tilde{j}}^{ij} - v_{\tilde{i}\tilde{j}}^{ji} \right) + V_{nuc}^A + V_{nuc}^B + \frac{1}{2} \sum_{I,J\notin AB} \frac{Z_I Z_J}{R_{IJ}} \end{aligned} \tag{A8}$$

We can now transform the above equation into the atomic orbital basis by introducing the spin density matrix $P_{\mu\nu}^{\sigma} = \sum_i C_{\mu i}^{\sigma} \tilde{C}_{i\nu}^{\sigma}$ and the total density matrix $D_{\mu\nu} = P_{\mu\nu}^{\alpha} + P_{\mu\nu}^{\beta}$, where $C_{\mu i}$ is the coefficient of basis function μ for orbital i, and $\tilde{C}_{\mu i}$ is the coefficient of the corresponding contravariant orbital.

$$\begin{aligned} E_0 &= \sum_{\mu\nu} \left(P_{\mu\nu}^{\alpha} + P_{\mu\nu}^{\beta} \right) \left(\hat{h}_0\right)_{\nu\mu} + \frac{1}{2} \sum_{\mu\nu\lambda\sigma} \left(P_{\mu\nu}^{\alpha} + P_{\mu\nu}^{\beta} \right) \left(\hat{J}_0\right)_{\nu\mu} \\ &\quad - \frac{1}{2} \sum_{\mu\nu\lambda\sigma} \left(P_{\mu\nu}^{\alpha} \left(\hat{K}_0\right)_{\nu\mu}^{\alpha} + P_{\mu\nu}^{\beta} \left(\hat{K}_0\right)_{\nu\mu}^{\beta} \right) + V_{nuc}^A + V_{nuc}^B + \frac{1}{2} \sum_{I,J\notin AB} \frac{Z_I Z_J}{R_{IJ}} \end{aligned} \tag{A9}$$

where we have introduced a shortened notation for the integrals, making use of the strict orbital localization:

$$\left(\hat{h}_0\right)_{\nu\mu} = \begin{cases} \left(\hat{h}_{AC}\right)_{\nu\mu} & \text{if } \mu \in A \\ \left(\hat{h}_{BC}\right)_{\nu\mu} & \text{if } \mu \in B \\ \left(\hat{h}\right)_{\nu\mu} & \text{if } \mu \in C \end{cases} \tag{A10}$$

$$\left(\hat{J}_0\right)_{\nu\mu} = \begin{cases} D_{\lambda\sigma} \left(\hat{v}_{AC}\right)_{\nu\sigma}^{\mu\lambda} & \text{if } \mu,\lambda \in A \\ D_{\lambda\sigma} \left(\hat{v}_{BC}\right)_{\nu\sigma}^{\mu\lambda} & \text{if } \mu,\lambda \in B \\ D_{\lambda\sigma} v_{\nu\sigma}^{\mu\lambda} & \text{if } \mu,\lambda \notin \{AB\} \end{cases} \tag{A11}$$

$$\left(\hat{K}_0\right)^{\sigma}_{\nu\mu} = \begin{cases} P^{\sigma}_{\lambda\sigma}\left(\hat{v}_{AC}\right)^{\lambda\mu}_{\nu\sigma} & \text{if } \mu,\lambda \in A \\ P^{\sigma}_{\lambda\sigma}\left(\hat{v}_{BC}\right)^{\lambda\mu}_{\nu\sigma} & \text{if } \mu,\lambda \in B \\ P^{\sigma}_{\lambda\sigma} v^{\lambda\mu}_{\nu\sigma} & \text{if } \mu,\lambda \notin \{AB\} \end{cases} \tag{A12}$$

Finally, we introduce the Fock matrix $\hat{F}^{\sigma}_0 = \hat{h}_0 + \hat{J}_0 - \hat{K}^{\sigma}_0$, and we retrieve the usual formulation for the Hartree-Fock energy:

$$E_0 = \frac{1}{2}\sum_{\mu\nu} P^{\alpha}_{\mu\nu}\left(\left(\hat{h}_0\right)_{\nu\mu} + \left(\hat{F}^{\alpha}_0\right)_{\nu\mu}\right) + \frac{1}{2}\sum_{\mu\nu} P^{\beta}_{\mu\nu}\left(\left(\hat{h}_0\right)_{\nu\mu} + \left(\hat{F}^{\beta}_0\right)_{\nu\mu}\right) + V^A_{nuc} + V^B_{nuc} + \frac{1}{2}\sum_{I,J\notin AB}\frac{Z_I Z_J}{R_{IJ}} \tag{A13}$$

From here, the reader is invited to follow the derivation presented for the zeroth-order wavefunction[35] or by Stoll[45]. After differentiation of the energy including the orbital localization constraints, the SCF equations for the zeroth-order wavefunction are obtained.

Appendix B: Remainder of the Hamiltonian

The remaining term in the Hamiltonian takes the form

$$\begin{aligned} \widehat{H}_{BE} = &\sum_{i\in A}\sum_{k}\left\langle \tilde{k}\left|\hat{h}_{\overline{AC}}\right|i\right\rangle k^{+}\tilde{i}^{-} + \sum_{i\in B}\sum_{k}\left\langle \tilde{k}\left|\hat{h}_{\overline{BC}}\right|i\right\rangle k^{+}\tilde{i}^{-} + \\ &+\sum_{k,l\in A}\sum_{i,j}\left\langle \tilde{i}\tilde{j}\left|\hat{v}_{\overline{AC}}\right|kl\right\rangle i^{+}j^{+}\tilde{l}^{-}\tilde{k}^{-} + \sum_{k,l\in B}\sum_{i,j}\left\langle \tilde{i}\tilde{j}\left|\hat{v}_{\overline{BC}}\right|kl\right\rangle i^{+}j^{+}\tilde{l}^{-}\tilde{k}^{-} \quad , \end{aligned} \tag{B1}$$

with one-electron operators

$$\hat{h}_{\overline{AC}} = \left(1-\hat{P}_{AC}\right)\left(\hat{V}_A + \hat{V}_C + \hat{T}\right) \tag{B2}$$

$$\hat{h}_{\overline{BC}} = \left(1-\hat{P}_{BC}\right)\left(\hat{V}_B + \hat{V}_C + \hat{T}\right) \tag{B3}$$

and the two-electron operators

$$\hat{v}_{\overline{AC}} = \frac{1}{2}\left(1-\hat{P}_{AC}\right)(1)\left(1-\hat{P}_{AC}\right)(2)\frac{1}{r_{12}} \tag{B4}$$

$$\hat{v}_{\overline{BC}} = \frac{1}{2}\left(1-\hat{P}_{BC}\right)(1)\left(1-\hat{P}_{BC}\right)(2)\frac{1}{r_{12}} \tag{B5}$$

Components in (B2) correspond to the description of pairs of fragments A and C, and B and C by the basis functions assigned to the orthogonal complement of those pairs of fragments. Since in the zeroth-order wavefunction any links between fragments A and B must be removed, there is no place for those terms in the zeroth-order Hamiltonian. On the other hand, those components do not correspond to any physical interaction between A and B, but are rather a basis set effect, similar to the "basis set extension error" term in Mayer's[41] Chemical Hamiltonian. Following Mayer, we treat it as such and discard it from the perturbation.

Appendix C: Components of the second-order energy corrections

To obtain the explicit form of the second-order energy correction (see Eq. 8), one needs to evaluate elements $\left\langle \tilde{\Psi}_{exc} \left| \widehat{W}_{AB} \right| \Psi^{(0)} \right\rangle$ and $\left\langle \tilde{\Psi}^{(0)} \left| \widehat{W}_{AB} \right| \Psi_{exc} \right\rangle$. Since in a single-determinant approximation only singly and doubly excited determinants will contribute to the second-order corrections, the only non-zero elements have the forms: $\left\langle \tilde{\Psi}^{(0)} \left| \widehat{W}_{AB} \right| \Psi_a^b \right\rangle$, $\left\langle \tilde{\Psi}_a^b \left| \widehat{W}_{AB} \right| \Psi^{(0)} \right\rangle$, $\left\langle \tilde{\Psi}_{ac}^{bd} \left| \widehat{W}_{AB} \right| \Psi^{(0)} \right\rangle$ and $\left\langle \tilde{\Psi}_0 \left| \widehat{W}_{AB} \right| \Psi_{ac}^{bd} \right\rangle$. The first element reads

$$\left\langle \hat{\Psi}^{(0)} \left| \widehat{W}_{AB} \right| \Psi_a^b \right\rangle = \sum_{i\in A}\sum_{k} \left\langle \tilde{k} \left| \hat{V}_B \right| i \right\rangle \left\langle \tilde{\Psi}^{(0)} \left| k^+\tilde{i}^- \right| \Psi_a^b \right\rangle + \sum_{i\in B}\sum_{k} \left\langle \tilde{k} \left| \hat{V}_B \right| i \right\rangle \left\langle \tilde{\Psi}^{(0)} \left| k^+\tilde{i}^- \right| \Psi_a^b \right\rangle + \sum_{k\in A}\sum_{l\in B}\sum_{ij} \langle \tilde{i}\tilde{j}|kl\rangle \left\langle \tilde{\Psi}^{(0)} \left| i^+j^+\tilde{l}^-\tilde{k}^- \right| \Psi_a^b \right\rangle \quad \text{(C1)}$$

Using the biorthogonality condition and the anticommutation rules, one obtains

$$\left\langle \tilde{\Psi}^{(0)} \left| \widehat{W}_{AB} \right| \Psi_a^b \right\rangle = \left\langle \tilde{a} \left| \hat{V}_B \right| b \right\rangle \delta_b^A + \left\langle \tilde{a} \left| \hat{V}_A \right| b \right\rangle \delta_b^B + \sum_{l\in B}^{occ} \left\langle \tilde{a}\tilde{l}||bl \right\rangle \delta_b^A + \sum_{k\in A}^{occ} \left\langle \tilde{a}\tilde{k}||bk \right\rangle \delta_b^B \quad \text{(C2)}$$

where

$$\delta_b^A = \begin{cases} 1, b \in A \\ 0, b \notin A \end{cases} \quad \text{(C3)}$$

Similarly,

$$\left\langle \tilde{\Psi}_a^b \left| \widehat{W}_{AB} \right| \Psi^{(0)} \right\rangle = \left\langle \tilde{b} \left| \hat{V}_B \right| a \right\rangle \delta_a^B + \left\langle \tilde{b} \left| \hat{V}_A \right| a \right\rangle \delta_a^A + \sum_{l\in B}^{occ} \left\langle \tilde{b}\tilde{l}||al \right\rangle \delta_a^A + \sum_{k\in A}^{occ} \left\langle \tilde{b}\tilde{k}||ak \right\rangle \delta_a^B \quad \text{(C4)}$$

The doubly excited elements read

$$\left\langle \tilde{\Psi}^{(0)} \left| \widehat{W}_{AB} \right| \Psi^{bd}_{ac} \right\rangle = \langle \tilde{a}\tilde{c}||bd\rangle \, \delta^A_b \delta^B_d + \langle \tilde{a}\tilde{c}||bd\rangle \, \delta^B_b \delta^A_d \tag{C5}$$

and

$$\left\langle \tilde{\Psi}^{bd}_{ac} \left| \widehat{W}_{AB} \right| \Psi^{(0)} \right\rangle = \left\langle \tilde{b}\tilde{d}||ac \right\rangle \delta^A_b \delta^B_c + \left\langle \tilde{b}\tilde{d}||ac \right\rangle \delta^B_a \delta^A_c \tag{C6}$$

The component constisting of the products of doubly-excited elements is identified, following Surjan *et al.*[42] as dispersion energy. The remainder, the induction energy in our interpretation, built of singly-excited elements, consists of "single-fragment" terms containing products $\delta^A_a \delta^A_b$, $\delta^B_a \delta^B_b$ and cross terms $\delta^A_a \delta^B_b$. The one-fragment terms are grouped into a component called polarization energy and the cross terms are identified as delocalization energy (also following Surjan *et al.*[42]).

REFERENCES

[1]Jeffrey, G. A. *An introduction to hydrogen bonding*, Vol. 12; Oxford University Press New York, 1997.

[2]Metrangolo, P.; Meyer, F.; Pilati, T.; Resnati, G.; Terraneo, G. *Angew. Chem. Int. Ed.* **2008**, *47*(33), 6114–6127.

[3]Hagemann, M.; Berger, R. J.; Hayes, S. A.; Stammler, H.-G.; Mitzel, N. *Chem. Eur. J.* **2008**, *14*(35), 11027–11038.

[4]Scherer, W.; Sirsch, P.; Shorokhov, D.; Tafipolsky, M.; McGrady, G. S.; Gullo, E. *Chem. Eur. J.* **2003**, *9*(24), 6057–6070.

[5]Dougherty, D. A.; Stauffer, D. A. *Science* **1990**, *250*(4987), 1558–1560.

[6]Schneider, H.-J. *Angew. Chem. Int. Ed.* **2009**, *48*(22), 3924–3977.

[7]Scheiner, S. *Noncovalent Forces;* Cham: Springer International Publishing, 2015.

[8]Raynal, M.; Ballester, P.; Vidal-Ferran, A.; van Leeuwen, P. W. *Chem. Soc. Rev.* **2014**, *43*(5), 1660–1733.

[9]McMullin, C. L.; Jover, J.; Harvey, J. N.; Fey, N. *Dalton Trans.* **2010**, *39*(45), 10833–10836.

[10]McMullin, C. L.; Fey, N.; Harvey, J. N. *Dalton Trans.* **2014**, *43*(36), 13545–13556.

[11]Ehrlich, S.; Bettinger, H. F.; Grimme, S. *Angew. Chem. Int. Ed.* **2013**, *52*(41), 10892–10895.

[12]Scheraga, H. A. *The Proteins Composition, Structure, and Function* **2012**, *1*, 477.

[13]Brillante, A.; Bilotti, I.; Della Valle, R. G.; Venuti, E.; Girlando, A. *CrystEngComm* **2008**, *10*(8), 937–946.

[14]Ferella, L.; Rosato, A.; Turano, P.; Plavec, J. *What Can be Learned about the Structure and Dynamics of Biomolecules from NMR;* Wiley Online Library, 2012.

[15]Lin, S.; Chen, J.-L.; Huang, L.-S.; Lin, H.-W. *Curr. Proteomics* **2005**, *2*(1), 55–81.

[16]Munshi, P.; Guru Row, T. N. *J. Phys. Chem. A* **2005**, *109*(4), 659–672.

[17]van Oijen, A. M. *Curr. Opin. Biotechnol.* **2011**, *22*(1), 75–80.

[18]Johnson, E. R.; Keinan, S.; Mori-Sanchez, P.; Contreras-Garcia, J.; Cohen, A. J.; Yang, W. *J. Am. Chem. Soc.* **2010**, *132*(18), 6498–6506.

[19]De Silva, P.; Corminboeuf, C. *J. Chem. Theory Comput.* **2014**, *10*(9), 3745–3756.

[20]Bader, R. F. *Chem. Rev.* **1991**, *91*(5), 893–928.

[21]Kitaura, K.; Morokuma, K. *Int. J. Quantum Chem.* **1976**, *10*(2), 325–340.

[22]Schuetz, M.; Rauhut, G.; Werner, H.-J. *J. Phys. Chem. A* **1998**, *102*(29), 5997–6003.

[23]Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. *J. Chem. Phys.* **1996**, *105*(5), 1968–1986.

[24]Addicoat, M. A.; Collins, M. A. *J. Chem. Phys.* **2009**, *131*(10), 104103.

[25]Mochizuki, Y.; Fukuzawa, K.; Kato, A.; Tanaka, S.; Kitaura, K.; Nakano, T. *Chem. Phys. Lett.* **2005**, *410*(4-6), 247 – 253.

[26]Mo, Y.; Gao, J.; Peyerimhoff, S. D. *J. Chem. Phys.* **2000**, *112*(13), 5530–5538.

[27]Khaliullin, R. Z.; Head-Gordon, M.; Bell, A. T. *J. Chem. Phys.* **2006**, *124*(20), 204105.

[28]Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chem. Rev.* **1994**, *94*(7), 1887–1930.

[29]Hohenstein, E. G.; Sherrill, C. D. *J. Chem. Phys.* **2010**, *133*(10), 104107.

[30]Hohenstein, E. G.; Jaeger, H. M.; Carrell, E. J.; Tschumper, G. S.; Sherrill, C. D. *J. Chem. Theory Comput.* **2011**, *7*(9), 2842–2851.

[31]Hohenstein, E. G.; Parrish, R. M.; Sherrill, C. D.; Turney, J. M.; Schaefer III, H. F. *J. Chem. Phys.* **2011**, *135*(17), 174107.

[32]Dodziuk, H.; Korona, T.; Lomba, E.; Bores, C. *J. Chem. Theory Comput.* **2012**, *8*(11), 4546–4555.

[33]Parrish, R. M.; Sherrill, C. D. *J. Chem. Phys.* **2014**, *141*(4), 044115.

[34]Parrish, R. M.; Parker, T. M.; Sherrill, C. D. *J. Chem. Theory Comput.* **2014**, *10*(10), 4417–4431.

[35]Gonthier, J. F.; Corminboeuf, C. *J. Chem. Phys.* **2014**, *140*(15), 154107.

[36]Parrish, R. M.; Gonthier, J. F.; Corminbœuf, C.; Sherrill, C. D. *J. Chem. Phys.* **2015**, *143*(5), 051103.

[37]Su, P.; Chen, Z.; Wu, W. *Chem. Phys. Lett.* **2015**, *635*, 250–256.

[38]Hobza, P.; Zahradník, R.; Müller-Dethlefs, K. *Collect. Czech. Chem. Commun.* **2006**, *71*(4), 443–531.

[39]Černỳ, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2007**, *9*(39), 5291–5303.

[40]Wagner, J. P.; Schreiner, P. R. *Angew. Chem. Int. Ed.* **2015**, page in print.

[41]Mayer, I. *Int. J. Quantum Chem.* **1983**, *23*(2), 341–363.

[42]Surján, P. R.; Mayer, I.; Lukovits, I. *Chem. Phys. Lett.* **1985**, *119*(6), 538–542.

[43]Mayer, I. *Int. J. Quantum Chem.* **1998**, *70*(1), 41–63.

[44]Mayer, I.; Vibok, A. *Mol. Phys.* **1997**, *92*(3), 503–510.

[45]Stoll, H.; Wagenblast, G.; Preuβ, H. *Theor. Chem. Acc.* **1980**, *57*(2), 169–178.

[46]Smits, G. F.; Altona, C. *Theor. Chem. Acc.* **1985**, *67*(6), 461–475.

[47]Gianinetti, E.; Raimondi, M.; Tornaghi, E. *Int. J. Quantum Chem.* **1996**, *60*(1), 157–166.

[48]Mo, Y.; Peyerimhoff, S. D. *J. Chem. Phys.* **1998**, *109*(5), 1687–1697.

[49]Sironi, M.; Famulari, A. *Theor. Chem. Acc.* **2000**, *103*(5), 417–422.

[50]Nagata, T.; Takahashi, O.; Saito, K.; Iwata, S. *J. Chem. Phys.* **2001**, *115*(8), 3553–3560.

[51]Kaplan, I. G. *Theory of molecular interactions;* Elsevier Science Ltd, 1986.

[52]Wilson, S. *Electron correlation in molecules;* Courier Corporation, 2014.

[53]Vibók, Á.; Mayer, I. *Acta Phys. Hung.* **1990**, *68*(3-4), 241–251.

[54]Gill, P. M.; Radom, L. *Chem. Phys. Lett.* **1986**, *132*(1), 16–22.

[55]Beran, G. J.; Gwaltney, S. R.; Head-Gordon, M. *Physical Chemistry Chemical Physics* **2003**, *5*(12), 2488–2493.

[56]Su, P.; Li, H. *J. Chem. Phys.* **2009**, *131*(1), 014102.

[57]Azar, R. J.; Horn, P. R.; Sundstrom, E. J.; Head-Gordon, M. *J. Chem. Phys.* **2013**, *138*(8), 084102.

[58]Mayer, I. *Chem. Phys. Lett.* **2000**, *332*(3), 381–388.

[59]Steinmann, S. N.; Corminboeuf, C.; Wu, W.; Mo, Y. *J. Phys. Chem. A* **2011**, *115*(21), 5467–5477.

[60]Molpro, version 2010.1, a package of ab initio programs. Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M.; others. **2010**.

[61]See supplemental material at [URL will be inserted by AIP] for results of intra-SAPT computations for rare gas dimer systems.

[62]Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*(5), 2257–2261.

[63]Brassell, S.; Eglinton, G.; Maxwell, J.; Philp, R. *Aquatic pollutants: Transformation and biological effects* **1978**, pages 69–86.

[64]Eichmann, R.; Neuling, P.; Ketseridis, G.; Hahn, J.; Jaenicke, R.; Junge, C. *Atmospheric Environment (1967)* **1979**, *13*(5), 587–599.

[65]Lúttschwager, N. O. B.; Wassermann, T. N.; Mata, R. A.; Suhm, M. A. *Angew. Chem. Int. Ed.* **2013**, *52*(1), 463–466.

[66]Byrd, J. N.; Bartlett, R. J.; Montgomery Jr, J. A. *J. Phys. Chem. A* **2014**, *118*(9), 1706–1712.

[67]Liakos, D. G.; Neese, F. *J. Chem. Theory Comput.* **2015**, *11*(5), 2137–2143.

[68]Echeverría, J.; Aullón, G.; Danovich, D.; Shaik, S.; Alvarez, S. *Nature Chemistry* **2011**, *3*(4), 323–330.

[69]Wagner, J. P.; Schreiner, P. R. *J. Chem. Theory Comput.* **2014**, *10*(3), 1353–1358.

[70]Hariharan, P. C.; Pople, J. A. *Theor. Chem. Acc.* **1973**, *28*(3), 213–222.

[71]Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*(18), 3297–3305.

[72]Dunning Jr., T. H. *J. Chem. Phys.* **1989**, *90*(2), 1007–1023.

[73]Wu, W.; Liu, Y.; Zhu, D. *Chem. Soc. Rev.* **2010**, *39*, 1489–1502.

[74]Podeszwa, R.; Bukowski, R.; Szalewicz, K. *J. Phys. Chem. A* **2006**, *110*(34), 10345–10354.

[75]Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2003**, *107*(41), 8377–8379.

[76]Sinnokrot, M. O.; Sherrill, C. D. *J. Am. Chem. Soc.* **2004**, *126*(24), 7690–7697.

[77]Hsu, S.-M.; Lin, Y.-C.; Chang, J.-W.; Liu, Y.-H.; Lin, H.-C. *Angew. Chem. Int. Ed.* **2014**, *53*(7), 1921–1927.

[78]Tsuzuki, S.; Uchimaru, T.; Mikami, M. *J. Phys. Chem. A* **2006**, *110*(5), 2027–2033.

[79]Wheeler, S. E.; Houk, K. *J. Am. Chem. Soc.* **2008**, *130*(33), 10854–10855.

[80]Umeyama, H.; Morokuma, K. *J. Am. Chem. Soc.* **1977**, *99*(5), 1316–1332.

[81]Beck, J. F.; Mo, Y. *J. Comput. Chem.* **2007**, *28*(1), 455–466.

[82]Hoja, J.; Sax, A. F.; Szalewicz, K. *Chem. Eur. J.* **2014**, *20*(8), 2292–2300.

[83]Jablonski, M.; Kaczmarek, A.; Sadlej, A. J. *J. Phys. Chem. A* **2006**, *110*(37), 10890–10898.

[84]Lane, J. R.; Contreras-García, J.; Piquemal, J.-P.; Miller, B. J.; Kjaergaard, H. G. *J. Chem. Theory Comput.* **2013**, *9*(8), 3263–3266.

[85]Tsuzuki, S.; Luthi, H. P. *J. Chem. Phys.* **2001**, *114*(9), 3949–3957.

[86]Kharitonov, Y.; Khoshabova, E.; Rodnikova, M. *Bull. Acad. Sci. USSR Div. Chem. Sci. (Engl.Transl.)* **1990**, *39*(6), 1190–1199.

[87]Thomsen, D. L.; Axson, J. L.; Schrøder, S. D.; Lane, J. R.; Vaida, V.; Kjaergaard, H. G. *J. Phys. Chem. A* **2013**, *117*(40), 10260–10273.

[88]Grabowski, S. J. *Annu. Rep. Prog. Chem., Sect. C: Phys. Chem.* **2006**, *102*, 131–165.

[89]Cooper, D. L.; Karadakov, P. B. *Int. Rev. Phys. Chem.* **2009**, *28*(2), 169–206.

[90]Hadad, C. M.; Rablen, P. R.; Wiberg, K. B. *J. Org. Chem.* **1998**, *63*(24), 8668–8681.

[91]Khaliullin, R. Z.; Bell, A. T.; Head-Gordon, M. *Chem. Eur. J.* **2009**, *15*(4), 851–855.